

Amendments to the Specification:

Please replace the paragraph beginning at page 5, line 1, which starts with "Modifications to the Euclidean" with the following amended paragraph:

A1
[page, 5, lines 1-9] Modifications to the Euclidean geometric model and the Euclidean distance metric may be able to address some of these shortcomings. A. Hinneburg, C. Aggarwal, and D. Keim, "What is the nearest neighbor in high dimensional spaces?" in *Proceedings of the 26th VLDB Conference*, 2000, [[has]] have proposed a variation on the conventional definition of similarity search to address the problem of dependencies. The method of Hinneburg et al. uses a heuristic to project the data set onto a low-dimensional subspace whose dimensions are chosen based on the point on which the similarity search is being performed. Because this approach is grounded in Euclidean geometry, it still incorporates some inherent disadvantages of Euclidean approaches.

Please replace the paragraph beginning at page 7, line 7, which starts with "In another aspect of the invention," with the following amended paragraph:

A2
[page 7, lines 7-13] In another aspect of the invention, embodiments of the search system preferably also support use of a query language that enables a general query for all items associated with a desired set of one or more properties. The result for such a query is the set of such items. In terms of the query language function, if two items are in the collection of items, [[than]] then the distance between them, in accordance with the particular distance function described above, is the smallest number of items returned by any of the queries whose results include both items.

Please replace the paragraph beginning at page 10, line 14, which starts with "This distance metric accounts" with the following amended paragraph:

A3
[page 10, line 14 - page 11, line 2] This distance metric accounts for the similarity between items based not only on the common occurrence of properties, but also their frequency. In

AB addition, this distance metric is meaningful in part because it captures the ~~dependence~~ dependencies among properties in the data. Normalized Euclidean distance metrics may take the frequency of properties into account, but they consider each property independently. The distance metric of the present invention takes into account the frequencies of combinations of properties. For example, Lawyer, College Graduate, and High-School Dropout may all be frequently occurring properties, but the combination Lawyer + College Graduate is much more frequent than the combination Lawyer + High-School Dropout. Thus, two lawyers who both dropped out of high school would be considered more similar than two lawyers who both graduated from college. Such an observation can be made if the distance metric takes into account the dependence among properties. In general, not all of the properties in the data will be useful for similarity search. For example, two people who share February 29th as a birthday may be part of a select group, but it is unlikely that this commonality reveals any meaningful similarity. Hence, in certain embodiments of the present invention, only properties deemed meaningful for assessing similarity are taken into account by the similarity search method. Properties that are deemed irrelevant to the search can be ignored.
